Vision Research 51 (2011) 2121–2131

Contents lists available at SciVerse ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The context effect in face matching: Effects of feedback

Bozana Meinhardt-Injac [a,*], Malte Persike [b], Günter Meinhardt [b]

[a] Department of Psychology, Binzmühlestrasse 14/22, University of Zürich, 8050 Zürich, Switzerland
[b] Johannes Gutenberg Universität, FB02, Department of Psychology, Methods Section, Wall str. 3, D-55122 Mainz, Germany

ARTICLE INFO

Article history:
Received 26 October 2010
Received in revised form 18 July 2011
Available online 10 August 2011

Keywords:
Holistic face processing
Context effect
Feedback

ABSTRACT

Faces are perceived holistically, even when they are presented briefly (Hole, 1994; Richler, Mack, et al., 2009). Results obtained with a context congruency paradigm support dominance of holistic processing for brief timings, but indicate that larger viewing times enable observers to regulate contextual influences, and to use a feature selective focus (Meinhardt-Injac, Persike, & Meinhardt, 2010). Here we provide further evidence for this claim, and illuminate the role of feedback. With trial by trial feedback observers show poor performance in incongruent facial contexts at brief timings, but become quite effective in suppressing information that interferes with the correct judgements at larger viewing times. Without feedback they are still able to delimit the effects of conflicting contextual information, but are less effective. Adding further target features leads to moderate performance increase in incongruent contexts when there is no feedback, but to strong improvement when feedback is provided. These findings indicate that observers use opportunities of learning to replace holistic face perception by modes of active vision when sufficient temporal resources are available.

1. Introduction

Faces are perceived holistically. This is what makes face perception "special", since there is no other object category which shows a comparable interdependence of parts. The integrative nature of face perception was recognized not later than Tanaka's and Farah's work showing that facial features are not encoded and represented independently, but together form a "Gestalt" as the smallest perceptual unit (Tanaka & Farah, 1993; Tanaka & Sengco, 1997). In their experiments facial features were much better identified in the context of a whole face than in isolation. With other visual objects such a part-to-whole advantage could not be observed to similar degrees. Holistic effects in face perception are now well studied, employing several experimental paradigms (see Maurer, Le Grand and Mondloch (2002), for a review).

One of the most popular paradigms designed to prove interaction among face parts is the composite face paradigm (Goffaux & Rossion, 2006; Hole, 1994; Rossion & Boremanse, 2008; Young, Hellawell, & Hay, 1987). In its original form, face composites are formed by combining a lower and a upper half, both stemming from different persons. In the experiment two such composite faces are shown, and observers have to judge upper or lower halves. Since perception tends to fuse both halves into a new whole face it is difficult to attend just one, ignoring the other. When the halves are misaligned (shifted against each other) or faces are inverted perceptual fusion resolves, and observers are able to selectively attend just the halves to be compared. Young et al. (1987) demonstrated that judging the identity of celebrities is slower and more error prone with aligned compared to misaligned halves. Hole (1994) used unfamiliar faces, and found that correct responses were faster for inverted than for upright composite face stimuli. However, this result was obtained only for brief presentation times of 80 ms, while for longer exposure durations of 2 s there was no advantage for inverted stimuli. Quite long response times of about 1700 ms for long presentations, which were about 500 ms longer than for brief presentations, indicated that subjects engaged in serial scanning and part-wise matching of the two images, which were shown side by side on the display. Since 80 ms are not enough for a saccade, these part-based strategies were precluded. Therefore, the study of Hole (1994) is the first to demonstrate that perceptual fusion of face halves occurs for briefly flashed face images. His results suggest that holistic face perception requires only brief encoding times, and is a first and natural mode of face vision, which may be replaced by other, more feature selective strategies if permitted by temporal resources and viewing conditions.

Although the study of Hole (1994) gives first hints that featural modes of face perception require more encoding time to be enabled than holistic modes it lacks to provide conclusive evidence. Particularly, the simultaneous side by side presentation of faces, combined with rather long viewing times, may have encouraged back and forth scanning among the two face images.

* Corresponding author.
E-mail addresses: b.meinhardt@psychologie.uzh.ch (B. Meinhardt-Injac), persike@uni-mainz.de (M. Persike), meinharg@uni-mainz.de (G. Meinhardt).

0042-6989/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2011.08.004

Two recent studies explicitly addressed the temporal encoding requirements of holistic face perception (Meinhardt-Injac, Persike, & Meinhardt, 2010; Richler, Mack, et al., 2009). Both studies used variants of a congruency paradigm to assess holistic effects. In a congruency paradigm the face stimulus is subdivided into two complementary areas, one to be attended (''target''), the other not to be attended (''no-target'', ''context''). In *congruent* trials, both parts are same in a pair (same trial), or both are different (different trial). In *incongruent* trials only the target face parts are same, while the no-target parts are different (same trial), or vice versa (different trial). If observers' performance depends on target and no-target face parts they perform well in congruent trials, but worse in incongruent trials. If they are able to attend just target face parts and to ignore no-target face parts performance in congruent and incongruent trials is equal. Hence, the performance difference obtained in congruent and incongruent trials quantifies a ''congruency effect'', which reflects the degree to which no-target face parts influence performance. The congruency effect has been shown to be much larger for faces than for other object categories (Farah et al., 1998; Gauthier et al., 2003), and is not due to response interference, like Stroop effects (Richler, Mack, et al., 2009).

In the study of Richler, Mack, et al. (2009) a fine raster of presentation times spanning from brief spatiotemporal twinkle (17 ms) up to relaxed timings of 800 ms was used. The most brief timings allowed just for chance performance, while at relaxed timings performance was nearly perfect. Congruency effects emerged for brief presentations below 50 ms, and reached constant levels for face presentations of about 100 ms, and beyond. Authors concluded that holistic viewing is the natural mode of face perception, which is enabled immediately, being neither replaced by other perceptual modes, nor attenuated at larger viewing times, in contrast to the results of Hole (1994).

Meinhardt-Injac and colleagues also varied presentation times, and measured face matching performance ranging from chance level to saturation levels of performance. In contrast to the study of Richler and colleagues it was found that both inversion effects, as well as congruency effects, were strongest at brief timings, but declined for presentation times beyond 200 ms. As in the Richler et al. study performance rose fast in congruent trials. In incongruent trials, however, there was further steady improvement at longer presentation times, indicating that observers exploited temporal resources to confine the influence of incongruent face parts in order to optimize performance. This was not observed in the Richler et al. study, where performance resided at moderate levels of about one d' unit despite ample viewing time.

Although it is widely accepted that perception is generally holistic when upright faces are viewed the finding that observers are seemingly unable to reduce the influence of incongruent facial parts when they have enough time and the resources to resort to more active viewing strategies is somewhat surprising. In experiments on perceptual learning it was shown that stimulus driven processing modes dominate the early response, but reside only if (i) task difficulty is high and (ii) mechanisms of feedback and control are disabled by the conditions of the task (Ahissar & Hochstein, 1993, 1997, 2000, 2004). Top-down processes and higher level control guide learning by suppressing task-irrelevant features and permit focusing relevant ones (Ahissar & Hochstein, 1993; Dupuis-Roy & Gosselin, 2007; Fahle & Morgan, 1996; Roelfsema, van Ooyen, & Watanabe, 2010). Sharpening of focused attention on task-relevant features is a key mechanisms in learning. Studies using tasks with simultaneous multiple feature variation were able to show that learning occurs only for task-relevant cues, but not for task-irrelevant cues, suggesting that mere stimulation is not sufficient for learning (Ahissar & Hochstein, 1993; Meinhardt & Grabbe, 2002; Shiu & Pashler, 1992).[1] Therefore, it can be expected that, when there is more encoding time, mechanisms allowing for more detailed access to individual aspects of a face stimulus come into play, and control mechanisms start to regulate contextual influences in order to better comply with the feature specific instructions of the task.

The conditions for focusing task-relevant facial features strongly differ in the experiments of Richler, Mack, et al. (2009) and Meinhardt-Injac, Persike, and Meinhardt (2010). In the Richler et al. study subjects saw face composites in study-image and test-image, presented consecutively. *After* the study-face they were informed whether the upper or the lower halves of two faces had to be matched. To be successful in incongruent trials therefore required to encode the whole face at study, to attend only the cued half as good as possible at test, and to bring together the results of both coding strategies at the decision. In the experiments of Meinhardt-Injac and colleagues the internal and the external facial features were used as target features and context features. The observers were told by instruction which of both sets was target. Therefore, they could try to focus the target features in *both* face presentations of a trial.

Second, no feedback about correctness was supplied in the Richler et al. study, while in the study of Meinhardt-Injac et al. subjects received trial by trial feedback. Feedback has been shown to be an effective means to accelerate learning (Herzog & Fahle, 1997, 1999), since it sets markers to task-relevant feature variation. As shown by Ahissar and Hochstein (1997, 2004), such markers are important in situations where observers are highly uncertain about the correct response – usually when task difficulty is high, or in case of visual illusions. Feedback, or some easy stimulus instances, serve to bring learning mechanisms on the right track again, and difficult stimulus instances are resolved in the following trials (''Eureka-effect''). As demonstrated in the presidential illusion (Sinha & Poggio, 1996) it is hard to recognize that it is really Bill Clinton's face, but with Al Gore's hairs and ears. Since face perception is naturally governed by the holistic viewing mode we recognize identity only when we are *told* to take a second look, or have the images side-by-side. It is therefore likely that subjects in the Richler et al. study did not ever realize that they were wrong, and that they could have been successful only with a part based strategy.

In view of these differences of both studies one may ask for the minimum temporal encoding requirements that allow other than holistic perceptual modes to enter in face perception. With providing feature certainty and feedback Meinhardt-Injac and colleagues found that congruency effects declined at longer exposure durations of beyond 200 ms, indicating that mechanisms of context regulation enter. Since authors studied the modulating effects of external context features on internal target features, and vice versa, it is mandatory to explore whether evidence for context regulation and control can also be found for other target feature sets, indicating feature independent temporal encoding requirements for these mechanisms in face perception. Further, since feedback may be crucial to calibrate feature selective viewing strategies in incongruent trials, the temporal encoding requirements of congruency effects should be studied with providing feedback, and without. Doing this should reveal whether cognitive markers are necessary for context regulation when observers try to selectively attend distinct aspects of a face.

2. Methods

2.1. Experimental outline

Three experiments were designed within the framework of the context congruency paradigm. In all three experiments two

[1] But see Seitz and Watanabe (2003, 2005) for task-irrelevant learning outside the focus of feature-selective attention.

consecutive face instances were to be matched according to agreement in internal target features. In Exp. I only eyes varied same/different within the set of internal features. In Exp. II we varied eyes and eyebrows, and in Exp. III the whole inner part of a face. All three experiments were executed with and without trial by trial feedback. Subjects were told which features of a face were possible targets in order to induce target feature certainty. Different subjects participated in each experiment, and in each feedback condition. Context congruency and exposure duration were within subject factors.

2.1.1. Context congruency

For each experiment face stimuli were composites, consisting of target features and the remainder face, which formed the context. Hence, target features and context were complementary in constituting a face. Two types of facial contexts were used. Subjects viewed faces in *congruent contexts* (CC), where the two face stimuli of a trial were either completely identical (same-trial) or completely different (different-trial). Further, they saw face stimuli in *incongruent contexts* (IC), where the two faces were same in the target features, and different in the remainder face (same-trial), or different in the target features, but same in the remainder face (different-trial). Hence, in all three experiments the unattended feature set could change congruently, or incongruently with the target feature set, thus providing either congruent (CC) or incongruent (IC) contextual information. Fig. 1 illustrates stimulus instances of the "same" category in incongruent contexts. Since it is clear that congruency refers to the relationship of target and context features, we refer to its modulatory effects as "context effects" in the following.

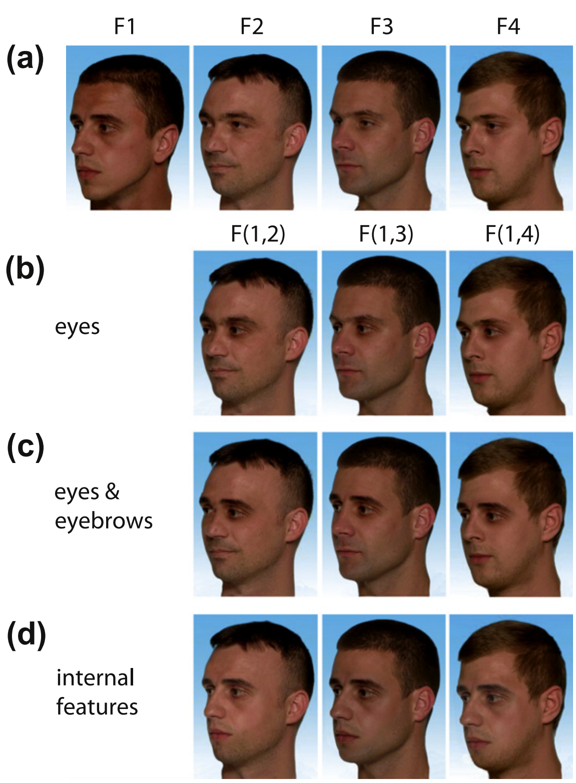

Fig. 1. The four original faces used for stimulus construction (left 3/4 view, a), and instances of same target features in incongruent external feature context for experiments I–III (b–d). In (b) the eyes of face 1 are mounted in faces 2–4. In (c) this is done with eyes and eyebrows, and in (d) with the whole set of internal features. Note that by moving from (b) to (d) the proportion of target features relative to context features increases.

2.1.2. Duration

Six exposure durations, $D = \{50, 133, 217, 333, 433, 633\}$ ms, were used in order to span a wide range of presentation times, ranging from brief timings, precluding saccades and serial scan, up to relaxed timings allowing detailed image scrutiny.

2.1.3. Design

The experimental design was a 3 (Experiment) $\times$ 2 (Feedback) $\times$ 2 (Context) $\times$ 6 (Duration) factorial plan with 72 conditions. Each experimental unit comprised 2 contexts and 6 exposure durations, which are the within subject conditions administered to a subject. In each of the six experimental units "same" and "different" trials were realized with 16 replications each, resulting in 384 trials. These were shuffled and assigned to a randomly ordered measurement list. Going through an experimental unit took about 20 min, which was done without pauses.

2.2. Stimuli

Photographs of four male face models were used as templates for stimulus construction. These were full-color 3/4 view photographs of the left face side captured in a photo studio under controlled lighting conditions, and using the same background for all photographs (see Meinhardt-Injac, Persike, and Meinhardt (2010), for more details). The original images were manipulated with Adobe Photoshop in order to construct sample stimuli with defined combinations of internal target features and remainder face surrounds. Internal features were cut out with comparable tracing lines and placed on the second (template) face, based on the position of the internal features. As illustrated in Fig. 1, both original and composite face are natural face images, and composite faces are not recognizable as such. For each of the three experiments we used a 4×4 facial feature matrix guiding composite stimulus construction, with line index referring to source face of internal features and column index referring to source face of external features, such that an entry $F(ij)$ denotes a face with internal target features of face i and surround of face j. For each of the 3 stimulus conditions (eyes/eyes and eyebrows/internal) this matrix allows us to form 24 different "same-internal" pairwise combinations, 24 possible different "same-external" pairwise combinations, 120 ways of forming "totally different" pairs and 16 ways to build "identical" faces. Hence, with 16 replications of trial types, choosing only 4 similar face models enables us to select pairwise combinations of face stimuli in all conditions of the design without repetition of the same stimulus pair.

2.3. Performance measures

The experiments aimed at determining the proportion of correct judgements as a function of exposure duration. In order to obtain proportion correct rates free of a possible response bias, they were calculated from correct "same" *and* correct "different" judgements. Since each trial type was realized in 16 replications, each proportion correct datum rests on $n = 32$ trials.

2.4. Subjects

In each of the six experimental units 21 subjects participated. They were undergraduate students, approximately 30% were male and 70% female in the six samples. All subjects had normal or corrected to normal vision. They had no former psychophysical experience, were paid and not informed about the purpose of the experiment. Each subject joined only one experimental group.

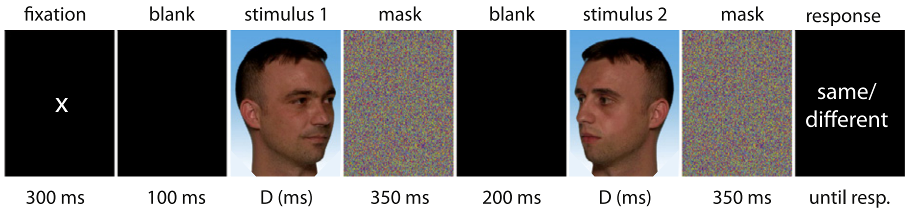

Fig. 2. Sequence of a trial.

2.5. Apparatus

The experiment was executed with Inquisit 2.0 runtime units. Patterns were displayed on NEC Spectra View 2090 TFT displays in 1280 × 1024 resolution at a refresh rate of 60 Hz. Screen mean luminance L_0 was 75 cd/m^2 at a michelson contrast of $(L_{max} - L_{min})/(L_{max} + L_{min}) = 0.98$, so the background was practically dark (about 1.4 cd/m^2, measured with a Cambridge Research Systems ColorCAL colorimeter). No gamma correction was used. The room was darkened so that the ambient illumination approximately matched the illumination on the screen to a fair degree. Patterns were viewed binocularly at a distance of 70 cm. Stimulus patterns and masks subtended 300 × 400 pixels (width × height), which corresponds to 12 × 15 cm of the screen, or 9.65° × 12° measured in degree of visual angle at 70 cm viewing distance. Subjects used a distance marker but no chin rest. They gave responses on an external numeric key-pad, and wore light headphones for acoustical feedback.

2.6. Procedure

A same/different forced choice task was used. In each experiment subject were informed that face pairs differed in the target features, but also in other respects. They were instructed to attend just the target features, and judge about their identity. The temporal order of events in a trial sequence was: fixation mark (300 ms) – blank (100 ms) – 1st stimulus frame (Duration) – mask (350 ms) – blank (200 ms) – 2nd stimulus frame (Duration) – mask (350 ms) – blank frame until response (see Fig. 2). Masking of the stimulus frames was done with spatial noise patterns with a grain resolution of 3 pixels. The presentation positions of each of the two face images were shifted by 20 pixels away from the center in random direction in order to preclude focusing the same image parts. First and second face images were leftward and rightward examples of the 3/4 view stimuli as a further means to obviate non-facial matching strategies. Acoustical trial by trial feedback about correctness, if provided, was given by brief headphone tone signals. Subjects were made familiar with the task by going through some randomly selected probe trials in order to ensure that the instruction was understood and could be put into practice.

3. Results

3.1. Proportion correct as a function of exposure duration

Fig. 3 shows proportion correct as a function of exposure duration for all experimental conditions. Data points indicate between subjects means, shown with their 95% confidence intervals. The proportion correct data were fitted with curves of exponential form

$$P(t) = 0.5 + b(1 - \exp(-a(t - t_0))) \qquad (1)$$

having b as the amplitude parameter, a as the scale parameter controling steepness, and t_0 as the location parameter. Parameters were estimated with a least squares criterion using the Levenberg–Marquardt algorithm. For all conditions the model fit was very good, with a ratio of explained to total variation larger than 95%. The parameters obtained from this procedure are listed in Table 1.

First inspection of the curves shows (i) monotonically rising performance with exposure duration in all conditions, (ii) generally increasing performance with increasing size of target feature set, (iii) strong performance difference for congruent (CC) and incongruent (IC) contexts (i.e. strong context effects), (iv) stronger context effects in the no-feedback condition than in the feedback condition.

For closer analysis the proportion correct data were fed into an ANOVA routine with "feature" and "feedback" as between subject factors and "context" and "exposure duration" as within subject factors. This analysis revealed significance of all main factors. Performance is better with increasing size of the target feature set ($F(2,120) = 31.44$, $p = .000$), better with than without feedback ($F(1,120) = 16.73$, $p = .000$), better in congruent than in incongruent contexts ($F(1,120) = 374.09$, $p = .000$), and better with increasing viewing times ($F(5,600) = 409.63$, $p = .000$). In order to obtain a detailed picture of the effects of feedback and target feature set size we analyzed the data separately for the first four viewing times, where performance still develops, and the longest two viewing times, where performance has settled to stable values in almost all conditions (see Fig. 4).[2] At the first four exposure durations there is no feedback effect in congruent contexts ($F(1,120) = 0.26$, $p = .619$), but in incongruent contexts ($F(1,120) = 15.99$, $p = .000$). There, performance is much better with feedback than without for matching internal features ($F(1,120) = 18.63$, $p = .000$), still better with feedback than without for matching eyes and eyebrows ($F(1,120) = 4.21$, $p = .042$), and same for both feedback conditions for matching eyes ($F(1,120) = 0.31$, $p = .576$). In order to reveal summation effects of target set size we calculated conditioned main effects of "feature" in all four feedback × context combinations, which is equivalent to calculating single ANOVA analyses for each of the four line plots shown in Fig. 4a. These analyses reveal that the effect of increasing target set size fails significance when there is no feedback in congruent contexts ($F(2,180) = 2.34$, $p = .105$, see open triangles in Fig. 4a) and also in incongruent contexts ($F(2,180) = 2.505$, $p = .090$, filled triangles in Fig. 4a). However, with feedback summation effects are strong in congruent ($F(2,180) = 15.49$, $p = .000$, open circles in Fig. 4a) and, particularly, in incongruent contexts ($F(2,180) = 21.59$, $p = .000$, filled circles). At the first four viewing times internal feature matching is done better than matching of eyes and eyebrows when there is feedback ($F(1,120) = 29.64$, $p = .000$), but not without ($F(1,120) = 2.94$, $p = .089$).

The picture changes when only the two longest exposure durations are analyzed, where performance has mostly settled to stable values (see Fig. 4b). For the effects of feedback the same pattern as obtained with the first four exposure durations results, with same performance in both feedback conditions for congruent contexts ($F(1,120) = 0.24$, $p = .626$), and better performance in the feedback condition than in the no feedback condition in incongruent

[2] A 3:3 data split would also roughly meet the saturation criterion, but we preferred the 4:2 split to include only saturation data, for all safety.

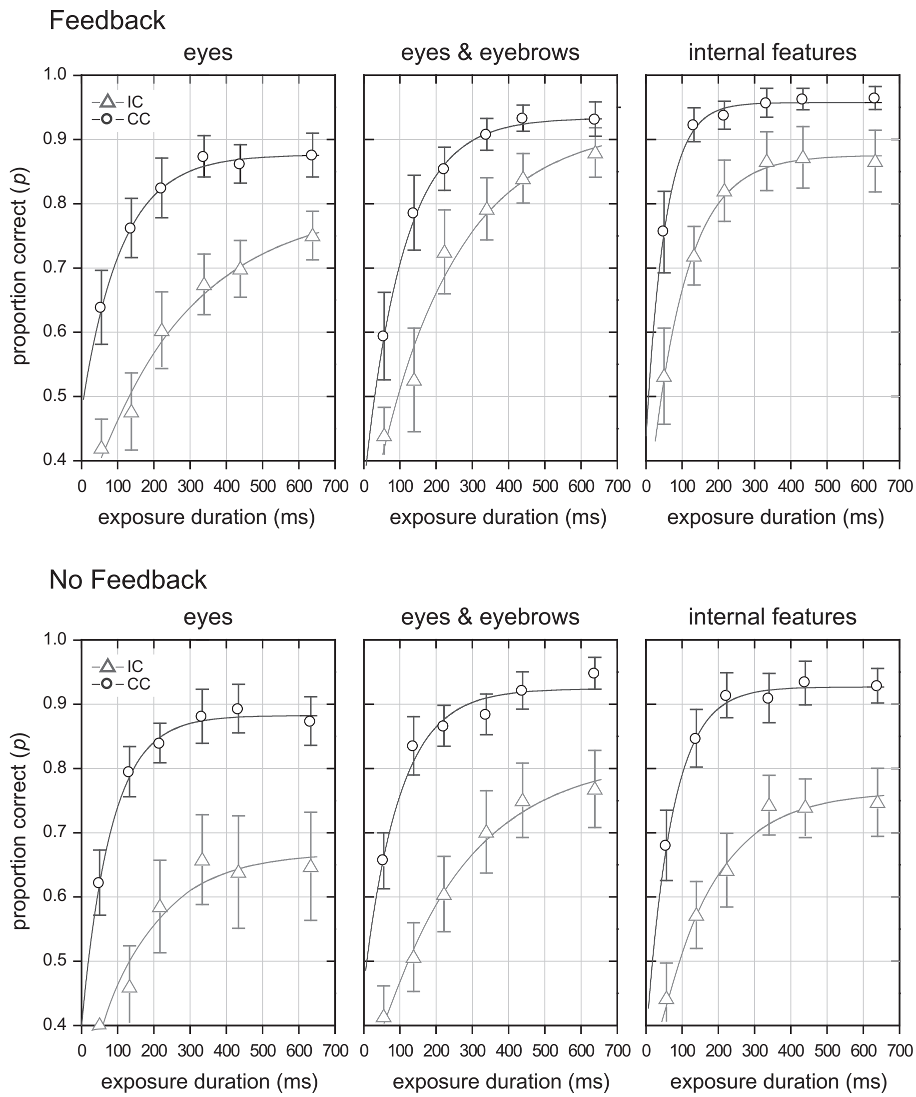

Fig. 3. Mean proportion correct rates as a function of exposure duration for face matching by eyes (left panel), eyes and eyebrows (mid panel), and the whole set of internal features (right panel) in congruent (black symbols) and incongruent contexts (gray symbols), with trial by trial feedback (upper panel) and without (lower panel). The smooth lines are exponential functions (1), with best fitting parameters for the least squares criterion. Error bars indicate 95% confidence limits based on the standard error of the mean for between subject variation.

Table 1
Parameters of the fitting curves for the exponential model (1).

Feedback	Feature	Context	b	a	t_0
NoFB	Eyes	CC	0.382	0.012	18.91
NoFB	Eyes	IC	0.169	0.006	133.14
NoFB	Eyes and eyebrows	CC	0.424	0.010	3.21
NoFB	Eyes and eyebrows	IC	0.318	0.004	115.26
NoFB	Eyes and internal	CC	0.427	0.014	11.29
NoFB	Eyes and internal	IC	0.264	0.007	82.60
FB	Eyes	CC	0.376	0.009	1.20
FB	Eyes	IC	0.299	0.004	125.33
FB	Eyes and eyebrows	CC	0.434	0.009	23.73
FB	Eyes and eyebrows	IC	0.423	0.005	88.51
FB	Eyes and internal	CC	0.457	0.019	6.61
FB	Eyes and internal	IC	0.375	0.010	42.55

contexts ($F(1,120) = 23.1$, $p = .000$). However, feature summations effects are different. Conditioned main effect analysis indicates significance of the feature factor for each of the four line plots shown in Fig. 4b, but pairwise comparisons reveal that it is only the deviation of performance with only eyes from the performance achieved with the other target set sizes that is the source of statistical significance. Pairwise comparisons among "eyes and eyebrows" and "internal features" are not significant in both contexts and feedback conditions.

Taken together, statistical analysis shows that (i) the effects of feedback are present only in incongruent contexts, and (ii) effects of increasing target feature set size must be distinguished with respect to exposure duration. For the first four exposure durations there are significant feature summation effects only in the feedback condition. There, performance increases monotonically with increasing target set size. At the two longest exposure durations there is feature summation in both feedback conditions. However, this effect merely reflects poorer performance with just eyes as the target features, while performance is same for attending eyes and eyebrows, and internal features. This results pattern is most pronounced for face matching in incongruent contexts (see Fig. 4b).[3]

[3] In the present study subjects were instructed to respond as accurate as possible. No instruction was given to also respond as quickly as possible. For all safety RT data were also recorded, but these are noisy, rather condition-unspecific, and reflect just a general decaying trend when plotted as a function of exposure duration. The RT data are provided as supplemental material for the electronic version of the article.

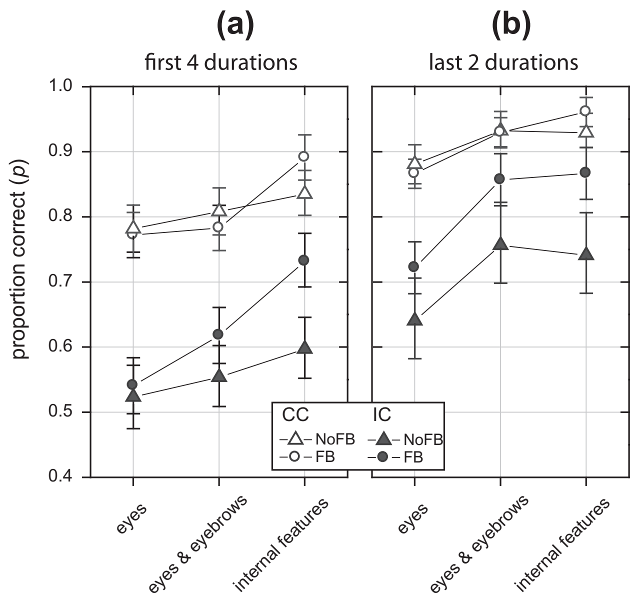

Fig. 4. Matching performance agglomerated across the first four exposure durations (left, a), and across only the two longest exposure durations (right, b). Open symbols represent means for congruent contexts, filled symbols for incongruent contexts, for matching with trial by trial feedback (circles), and without (triangles). Error bars indicate 95% confidence limits based on the standard error of the mean for between subject variation.

3.2. The context effect

We evaluated context effects on the level of individual performance, and analyzed the difference measure $\Delta p = p(CC) - p(IC)$ with ANOVA routines, having "feature" and "feedback" as between subject factors and "exposure duration" as within subject factor. Fig. 5 shows the mean differences for all three target feature set sizes and both feedback conditions as a function of exposure duration. Context effect rise to a maximum reached slightly above 100 ms, and then decline. With the exception of one case (matching eyes and eyebrows with feedback at the longest viewing time) context effects are significant, i.e. 0 is not within the confidence interval of the difference measure, Δp. The ANOVA routine indicates significance of exposure duration ($F(5,600) = 24.51$, $p = .000$) and feedback ($F(1,120) = 13.97$, $p = .000$), but not of feature ($F(2,120) = 3.38$, $p = .097$) on the main factors level, and no significant interactions. Testing the effect of feedback for each target feature set size shows stronger context effects without feedback than with feedback for matching internal features ($F(1,120) = 5.66$, $p = .019$) and for matching eyes and eyebrows ($F(1,120) = 7.12$, $p = .009$), but not for matching only eyes ($F(1,120) = 2.04$, $p = .156$). However, testing only at the two largest exposure durations shows stronger context effects without feedback also for matching only eyes ($F(1,120) = 6.45$, $p = .012$), in line with eye-inspection of the data shown in Fig. 5. The same test is also significant for matching eyes and eyebrows ($F(1,120) = 7.37$, $p = .008$), and for matching internal features ($F(1,120) = 6.14$, $p = .015$). At the two longest exposure durations the context effect is $\Delta p = 0.201$ without feedback and $\Delta p = 0.104$ with feedback, agglomerated across all three target set sizes. This means that at the largest viewing times there is a significant reduction of contextual influence of about 10% due to feedback.

3.3. Response bias

In order to account for possible response biases we applied d' data transformation, $d' = z(\mathrm{CR}) - z(\mathrm{Miss})$, and calculated the response criterion c on a standard axis with $d'/2$ as the new origin, as done previously (Richler, Bukach, & Gauthier, 2009). Here, "same" is assumed as target category, so positive values of c mean bias in favor of the "different" category, and negative values of c in favor of the "same" category. The c data were analyzed with the same ANOVA routine as the proportion correct data (see above). This analysis indicated significant effects of context ($F(1,120) = 57.01$, $p = .000$), exposure duration ($F(5,600) = 5.88$, $p = .000$), and context × exposure duration ($F(5,600) = 5.82$, $p = .000$). A data overview is shown in Fig. 6. The main effect of context reflects that judgements appear to be slightly biased towards the "same" category in congruent contexts ($\bar{c}_{CC} = -0.13$), while no overall bias is found in incongruent contexts ($\bar{c}_{IC} = -0.01$). The significant interaction of context and exposure duration reflects that the response criterion c is practically constant in congruent contexts, but, in incongruent contexts, it rises from negative values at the shortest exposure duration into a variation range about the zero expected value for all other exposure durations. Note that, at the level of the individual samples, significant deviations from

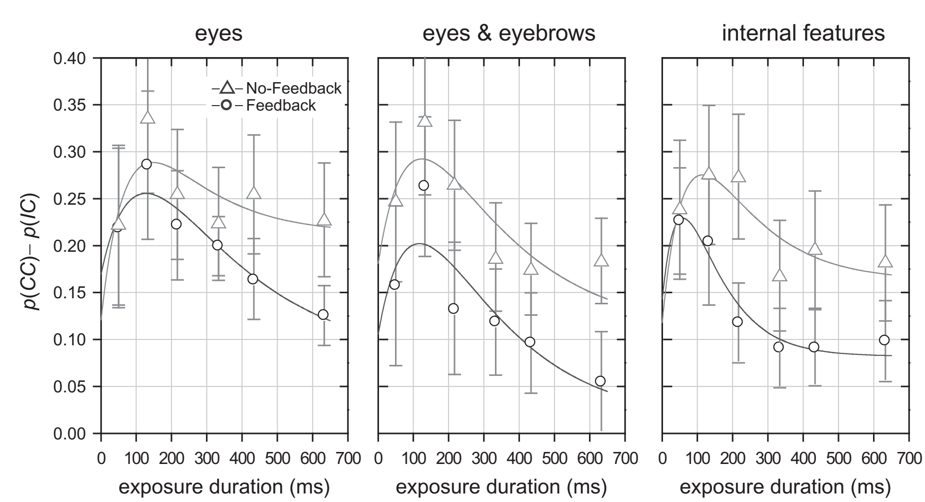

Fig. 5. Context effects, for face matching with trial by trial feedback (black symbols) and without feedback (gray symbols). The smooth lines are the differences of the corresponding fitting curves shown in Fig. 3. Error bars denote 95% confidence limits for the difference measure, Δp, based on its standard error for between subject variation. Note that a context effect is significant if 0 is outside its confidence interval.

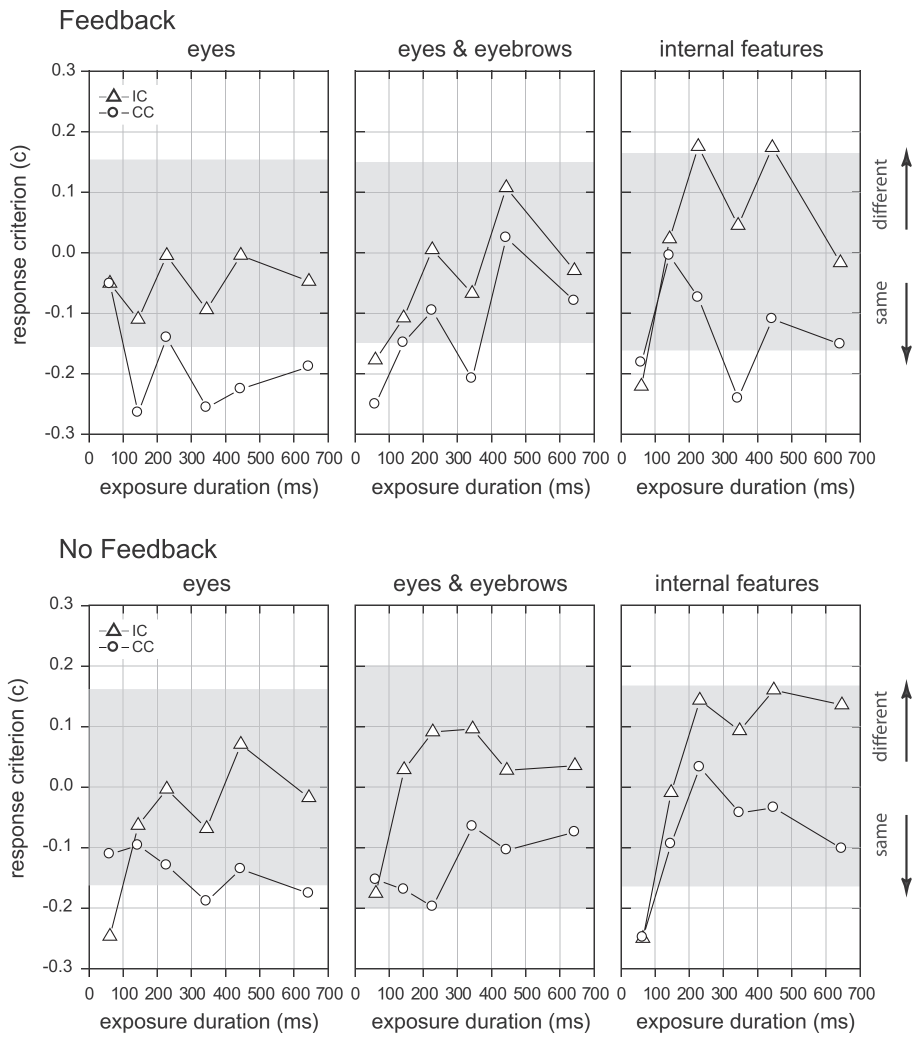

Fig. 6. Response criterion *c* (mean values) as a function of exposure duration for face matching by eyes (left panel), eyes and eyebrows (mid panel), and the whole set of internal features (right panel) in congruent (black symbols) and incongruent contexts (gray symbols), with trial by trial feedback (upper panel) and without (lower panel). A single 95% confidence interval for the zero expected value was calculated based on the samples shown in each of the six data panels (gray shaded area). Data points outside this range indicate a significant response bias.

the zero expected value hardly occur (see Fig. 6). Absolute values of *c* are 0.26 at the maximum.

4. Discussion

In three experiments we measured how face matching accuracy, as a function of exposure duration, is modulated by facial context, and how feedback mediates the effects of contextual modulation. In agreement with our former study (Meinhardt-Injac, Persike, & Meinhardt, 2010) we found that context effects are strong at brief timings, and decline with increasing exposure duration. Without feedback context effects are substantially larger than with feedback, and remain at strong levels of about 20% even for relaxed timings. Feedback was found to improve performance only in incongruent contexts. Both the dependency of the context effect on viewing time and the role of feedback in reducing the influence of facial context information that interferes with internal target feature information are strong support for development on the control axis in the temporal evolution of face perception, moving from holistic and automatic processing towards modes that allow to fade out incongruent contextual information, and to attend internal target features better focused, in line with instruction demands.

4.1. Context effects

Independent of feedback and internal feature target set size context effects take show a canonical dependency on exposure duration, peaking slightly beyond 100 ms and then continuously decline (see Fig. 5). In congruent contexts saturation levels of performance are reached at briefer viewing times than in incongruent contexts, implying that the declining sections of the context effect curves reflect that subjects exploit larger viewing times to better cope with incongruent contexts, while target features in congruent contexts can already be attended more easily, and with high accuracy. Same principal courses of context effects with feedback and without indicates that feedback is not necessary for regulation of contextual influence. However, with feedback contextual influence is reduced to much smaller levels at the same viewing times. Apparently, there are parallels to improvement in low level sensory learning, showing that task difficulty and feedback mutually contribute to optimization in sensory discrimination tasks which are handled in distal brain areas under higher level control (Ahissar & Hochstein, 1997; Herzog & Fahle, 1997; Meinhardt & Grabbe, 2002). At brief timings the task is difficult, and the observers have no leeway to probe strategies. With more temporal resources the task becomes easier, and they have capacity for ruling out ineffective strategies. Since feedback supplies immediate information about success a much more effective optimization process can be initialized within a 20 min session, particularly for the longer exposure durations.

The canonical dependency of contexts effects on encoding time for the three target set sizes indicates that holistic effects in face perception are overlayed and confined by mechanisms of active vision and strategy control, which enter when there is sufficient stimulus encoding time. Independent of feedback condition and target features set size context effects are off their peaks after about 200 ms presentation time, but strong at shorter exposure durations. Since presence or absence of feedback does not alter this, but affects just the overall level of contextual influence, the striking difference in the dependency of holistic effects on stimulus encoding time found in this study and the study of Richler, Mack, et al. (2009) is most likely due to the different target feature certainty conditions. In the Richler et al. study subjects were informed which face half was target briefly before test presentation, while at study they were uninformed about the target half. With target uncertainty subjects can perform well only when they encode the whole face at study. Further, randomly interleaved change of upper and lower face half as the targets at test let subjects no opportunities to build strategies of distracter influence control, since they did not know which half would be distracter, and which target. Hence, in the Richler et al. study holistic viewing is the best way of handling the task. However, as it can be concluded from the time course of holistic effects observed here, holistic viewing as a *strategy* is contingent with the particular certainty conditions in the Richler et al. study. As indicated by our results, with presentation times larger than 200 ms subjects are, in principle, able to resort to other, more effective modes of processing if focusing of task-relevant features is enabled by the certainty conditions of the task.

4.2. Response bias

Analysis of response bias reveals that subjects have a slight overall bias towards the "same" category in congruent contexts, independent of viewing time. In incongruent contexts, subjects are biased in favor of the "same" category at the briefest exposure durations, but are more likely to favor the "different" category, or show no bias at all, as exposure duration increases. However, on the z – standard scale these effects are rather small relative to performance, which increases to levels of beyond 2 when measured in d' units. Criterion shifts are mostly not significant when judged at the level of the individual time samples (see Fig. 6). Although the main effect of context reflects that, on a relative basis, subjects are more likely to respond "same" in congruent contexts compared to incongruent contexts, these bias effects are too small to indicate a substantial link of response category and context congruency in our paradigm.[4] Most important, the temporal changes of the response criterion show no systematic variation with the principal dependency of context effects on exposure duration: these are strong at brief timings, and then gradually decline, while, for criterion shifts, we find that there is a (small) bias towards the "same" category that stays independent of exposure duration in congruent contexts, and moves into a variation range about the neutral line in incongruent contexts. Therefore, a possible change of decision strategy with viewing time is no explanation for the declining context effects. In the study of Richler, Mack, et al. (2009) stronger shifts of response criteria in the range of 0.2–0.4 were observed at brief timings, which vanished for durations beyond 100 ms. Congruency effects were shown to develop beginning with brief timings below 50 ms, reaching their maximum shortly beyond 100 ms, and residing until the longest exposure durations. Therefore, it was found that response bias and congruency effects are *negatively correlated*. Authors did not comment on this finding, and did not try to explain it in terms of their paradigm. Apparently, there is a link of response bias and congruency effect in the paradigm used by Richler and colleagues, but the fact that response bias has vanished as congruency effects are at maximum implies that congruency effects cannot be due to response bias. Instead, one may conclude that *absence* of response bias favors congruency effects among attended and unattended face halves. These results contradict authors' recent claim that congruency effects may be strongly related to shifts in decision criteria (Richler et al., 2008, p. 341). As shown, such a link does not exist in our paradigm.

4.3. Target feature summation

Former work on feature summation effects in face perception has shown that these occur with spatially close target features, but not with distant ones (Thomas, 2001). Further, summation effects are seemingly bound to the upright orientation (Farivar & Chaudhuri, 2003). In agreement with Thomas' findings we find that adding eyebrows to eyes leads to strong increase of maximum performance, but adding further target features which are remote from the eyes/eyebrows region does not increase maximum performance any more (see Figs. 3 and 4b). Same maximum performance with eyes and eyebrows and the whole set of internal features as the target features indicates that the eyes/eyebrows region was mainly attended in our experiments. Experiments designed to measure the relative importance of different facial features in face recognition have consistently shown dominance of the eyes/eyebrows region, followed by the mouth and then the nose (Davies, Ellis, & Shepherd, 1977; Fraser, Craig, & Parker, 1990; Haig, 1986). Monitoring the scan paths of subjects also confirms that this region is most frequently attended (Henderson et al., 2001; Williams and Henderson, 2007). Eyes and eyebrows are most salient (Barton, Keenan, & Bass, 2001), and observers tend to keep this region in focus when performing an individual face discrimination task (Haig, 1985; Sekuler et al., 2004).

However, we observe same performance with eyes and eyebrows and the whole set of internal features as the targets only at the two longest viewing times. With limited temporal resources observers perform better with the whole set of internal features (see Fig. 4). This indicates that same performance with eyes and eyebrows and the whole set of internal features is the result of a viewing strategy, which is fully developed at relaxed viewing times. Indeed, trying to focus the eyes/eyebrows region and to keep the face identity cues extracted from this region free of contextual influence is a possible viewing strategy that consistently explains our results. At brief presentation times exactly focusing eyes and eyebrows is difficult due to position jitter and randomly swapping 3/4 views. When eyes and eyes and eyebrows are the target feature set the direct environment of the eyes/eyebrows region varies congruently and incongruently with the target features, while, with the whole set of internal features as the target features the direct neighborhood of the eyes/eyebrow region always varies congruently. Therefore, performance should be best for matching the whole set of internal features at brief timings. At longer viewing times observers are able to correct their focus, and are much better able to concentrate on just the eyes/eyebrows region. With eyes and eyebrows and internal features as the target feature set there is only congruent variation of features when the focus is well adjusted on the eyes/eyebrows region, leading to good accuracy levels when a narrow focus can be held. With only eyes as the target features congruently and incongruently varying eyebrows are still

[4] Criterion shifts in the range of [−0.2, 0.2] are more or less normal in forced choice experiments, and are negligible particularly at large d' values of 2 and beyond.

disturbing, so performance stays limited with only eyes. The development of a viewing strategy that keeps best track on the eyes/eyebrows region despite varying positions and swapping 3/4 views explains the scheme of results obtained for the effects of feature summation and feedback (see Fig. 4).

Even if such a strategy can be build during the course of the experiment with the aid of feedback feature summation effects in congruent contexts and the performance discrepancy of incongruent and congruent contexts show that global influences remain. Note that the best strategy in the three experiments would be to use a very narrow focus in Exp. I, a larger one in Exp. II, and the largest in Exp. III. The observation of small but monotonically increasing feature summation effects in congruent contexts (see open symbols in Fig. 4) confirms that subjects indeed preselected spatial foci of increasing size in the three experiments. Since, in congruent contexts, trials were identical in all three experiments and a wider focus serves to exploit target redundancy, better performance with increasing target set size can only be due to using larger spatial foci. However, the absence of feedback effects in congruent contexts and the *much* larger amplitude of feature summation effects in incongruent contexts indicates that strategy development mostly concentrated on handling incongruent spatial context information by learning incongruent context suppression and isolating target features from the surround.

4.4. Target feature certainty

In a recent debate about the mechanisms of holistic face processing the role of blocked presentation of change type conditions was critically discussed (see Riesenhuber & Wolff, 2009; Rossion, 2008). It was argued that only randomly interleaved presentation of relational and featural face image manipulations maintains holistic viewing strategies, while a-priori knowledge of observers that there will be just featural manipulations triggers analytical, local feature based modes of face vision. This was concluded from sequential effects in block mode presentation: if subjects judge trials with feature displacements first and trials with feature replacements thereafter performance is equal in both conditions. If the order is reversed, they are better with "featural" trials, and perform relatively poor with "configural" trials (Riesenhuber and Wolff, 2009; Riesenhuber et al., 2004).[5]

However, effects of inversion were observed for both blocking sequences. Further, Yovel and Kanwisher (2004) found same inversion costs for randomly interleaved presentation and blocking, having the subjects informed about the type of image manipulation. Also in this study and in the study of Meinhardt-Injac, Persike, and Meinhardt (2010) subjects had a-priori knowledge about the target feature set. Nevertheless, context effects are large – even at the longest viewing times there is 20% performance reduction due to incongruently varying contexts without feedback, and about 10% performance reduction when subjects are given better learning opportunities by providing feedback. As reported in the Meinhardt-Injac et al. study effects of inversion show, in principle, the same dependency on exposure duration than the effects of context, being strong at brief timings and then decline. As argued recently by reviewing a wide range of studies on the effects of inversion there is strong canonical evidence that the main effect of inversion is disruption of holistic face perception, with the consequence that assessment of long-range metric distances in face stimuli is particularly impaired (Rossion, 2008; Rossion & Boremanse, 2008; Sekunova & Barton, 2008).

As argued above, the striking difference in the temporal dependency of context congruency effects of this study and the Richler et al. study is most likely due to the different target feature certainty conditions. This is substantiated by recent findings about the time course of the inversion effect for featural and relational image manipulations. Meinhardt-Injac, Persike, and Meinhardt (2011) replicated the seminal study of Goffaux and Rossion (2007) on the differential effects of inversion for horizontal and vertical relational image manipulations, but added variation of exposure duration. For feature replacement and vertical displacement they found temporal dependencies of inversion effects that closely resembled the temporal dependencies of congruency effects in the Richler et al. study, emerging for briefly flashed images and then developing up to a plateau reached after about 200 ms exposure duration. For horizontal relational manipulations a small inversion effect, which emerged at larger presentation times, was found. In this study subjects received trial-by-trial feedback about correctness, as in the present study, but the three types of image manipulations were presented randomly interleaved, and without a cue. Hence, there was no target feature certainty, but feedback. Viewing the results of present studies together indicates that, without target feature certainty, correlates of holistic viewing arise for rapidly flashed face images, and reside over the whole spectrum of viewing times, while, with target feature certainty, they also arise for brief presentations, but decline for presentation times beyond 200 ms. This suggests that a possible replacement of early holistic and automatic processing modes by modes allowing for context regulation and sharpening the focus on task-relevant features is bound to the certainty conditions of the experiment. Feedback, however, is not necessary for feature selective processing to come into play.

4.5. The "microgenesis" of face perception

In the early days of face perception research Sergent (1986) proposed that the temporal evolution of face perception is organized along a coarse-to-fine continuum of facial feature resolution, with face outline and the major 1st order feature organization being rapidly present, and fine detail information of features and skin texture adding later. For all moments in time it is assumed that the facial representation is holistic, and tightly integrated, but at different resolution levels. Support for the temporal precedence of large scale information over finer scale information comes from studies on the role of spatial frequency content in face perception. Differential effects of spatial frequency filtering on matching performance with featural and relational face manipulations indicate that configural information is mostly retained after low-pass filtering, while featural information is impaired (Goffaux & Rossion, 2006; Goffaux, 2008, 2009). Low frequency information is processed faster than high frequency information in distal brain areas, and is handled by cells with large scale integration which are localized in the periphery of the visual field. High frequency information is processed slower, and serves fine detail analysis in the center of the visual field (Enroth-Cugell & Robson, 1966; Roufs, 1972; Roufs & Blommaert, 1981; Schyns & Oliva, 1994; Victor & Shapley, 1979).

Generally, if events require longer exposure duration this does not *imply* that they happen later in the stream of time. However, a possible source for the declining influence of context with viewing time is that fine scale information about internal facial features cannot develop to sufficient degrees when the retinal image and its representation in distal areas is curtailed by masking after short times. Feature specific matches, particularly for the eyes/eyebrows region, which contains fine details that make the difference, depend on good resolution levels of the facial representation. External context is fully present at very brief timings, at least at the level of global shape, and is capable to exert strong modulatory influence on internal features (Meinhardt-Injac, Persike, & Meinhardt, 2010). In line with this, the fact that for eyes and eyes and eyebrows performance is below chance at the briefest exposure duration in incongruent contexts

[5] Note that this claim was criticized as inconsistent with the results of other studies (see Rossion, 2008).

(see Fig. 3) suggests presence of fast and overruling external feature context.

The second assumption of Sergent (1986) besides the temporal precedence of large scale information is that processing is holistic starting with the first moments in time. Complementary to psychophysical evidence from rapid presentation and masking there is evidence from electrophysiological correlates of holistic viewing indicating that the initial face representations are holistic, and not part-based. Jacques and Rossion (2009) recorded ERPs generated in the right occipito-temporal cortex, and found that the face sensitive N170 (Bentin et al., 1996; Eimer, 2000) responds sensitive to the composite face illusion. Irrespective of true change of the top face half or just perceived change of top face half due to changing the lower face half there was a strong deflection of the N170 at about 160 ms. No effect was found for misaligned halves. In a follow-up study (Kuefner et al., 2010) authors controlled for possible confounds with decisional bias, since the composite effect was concluded from an increased rate in judging the upper halves as "different" although they were same. Again it was found that the N170 was modulated by true or illusory change of the top face half. However, the task was easy, and subjects did practically not make mistakes in the behavioral task, signaling top half changes only for true, but not for illusory changes. Corresponding to the behavioral responses there were electrophysiological correlates of later decisional components, as measured by lateralized readiness potentials starting between 310 and 330 ms, only for true top half changes. These results indicate that the perceptual basis of holistic viewing is early, being fully developed within the first 200 ms of processing. The first 200 ms of processing are known to be critical for encoding facial information, both at global and fine resolution levels. Face sensitive cells of the macaque inferior temporal cortex respond particularly in the interval of 100–200 ms after presentation (Tovee et al., 1993). Sugase and colleagues (Sugase et al., 1999), who recorded from the same area, found that global categorial information about faces was encoded in the earliest part of the responses, rising rapidly, peaking at 117 ms, and falling off before 200 ms. Information about face identity and emotional expression was found to be encoded with a latency of about 51 ms after the onset of global information transfer, peaking at 165 ms.

As indicated by our results, mechanisms of context regulation and sharpening a task-relevant featural focus do require larger encoding times than the mechanisms of holistic face processing. This is in contrast with low level visual tasks, such as texture segregation and pop-out detection. There, it could be shown that higher level control and learning concerns processing even at the most brief timings of about 20 ms (Ahissar & Hochstein, 1993; Karni & Sagi, 1991). However, whether facial feature selective learning can affect the earliest encoding stages of face perception remains to be proven by experiments explicitly dedicated to perceptual learning of this special object category.

Acknowledgments

This study was supported by the Swiss National Science Foundation Grant No. PMCD1-114398/1 and Stieffel-Zangger Nachwuchsförderkredit of Zürich University given to Bozana Meinhardt-Injac.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.visres.2011.08.004.

References

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Science, USA, 90*, 5718–5722.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 378*, 401–406.

Ahissar, M., & Hochstein, S. (2000). The spread of attention and learning in feature search: Effects of target distribution and task difficulty. *Vision Research, 40*, 1349–1364.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Science, 8*, 329–334.

Barton, J. J., Keenan, P. J., & Bass, T. (2001). Discrimination of spatial relations and features in faces: Effect of inversion and viewing duration. *British Journal of Psychology, 92*, 527–549.

Bentin, S., McCarthy, G., Perez, E., Puce, A., & Allison, T. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8*, 551–565.

Davies, G., Ellis, H., & Shepherd, J. (1977). Cue saliency in faces as assessed by the 'Photofit' technique. *Perception, 6*, 263–269.

Dupuis-Roy, N., & Gosselin, F. (2007). Perceptual learning without signal. *Vision Research, 47*, 349–356.

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology, 111*, 694–705.

Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology (London), 240*, 421–456.

Fahle, M., & Morgan, M. (1996). No transfer of perceptual learning between similar stimuli in the same retinal position. *Current Biology, 6*, 292–297.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. W. (1998). What is special about face perception? *Psychological Review, 105*, 482–498.

Farivar, R., & Chaudhuri, A. (2003). I can't see your eyes well 'cause your nose is too short. In R. Alterman, & D. Kirsh (Eds.), Proceedings of the twenty-fifth annual conference of the Cognitive Science Society, (pp. 378–383). Boston, Massachusetts, USA.

Fraser, I. H., Craig, G. L., & Parker, D. M. (1990). Reaction time measures of features saliency in schematic faces. *Perception, 19*, 661–673.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience, 6*, 428–432.

Goffaux, V. (2008). The horizontal and vertical relations in upright faces are transmitted by different spatial frequency ranges. *Acta Psychologica, 128*, 119–126.

Goffaux, V. (2009). Spatial interactions in upright and inverted faces: Re-exploration of spatial scale influence. *Vision Research, 49*, 774–781.

Goffaux, V., & Rossion, B. (2006). Faces are spatial holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1023–1039.

Goffaux, V., & Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 995–1002.

Haig, N. D. (1985). How faces differ – A new comparative technique. *Perception, 14*, 601–615.

Haig, N. D. (1986). Exploring recognition with interchanged facial features. *Perception, 15*, 235–247.

Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research, 37*, 2133–2141.

Herzog, M. H., & Fahle, M. (1999). Effects of biased feedback on learning and deciding in a vernier discrimination task. *Vision Research*, 39 4232–4243.

Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception, 23*, 65–74.

Henderson, J. M., Falk, R. J., Minut, S., Dyer, F. C., & Mahadevan, S. (2001). Gaze control for face learning and recognition by humans and machines. In T. Shipley & P. Kellman (Eds.), *From fragments to objects: Segmentation processes in vision* (pp. 463–482). Amsterdam: Elsevier.

Jacques, C., & Rossion, B. (2009). The initial representation of individual faces in the right occipito-temporal cortex is holistic: Electrophysiological evidence from the composite face illusion. *Journal of Vision, 9*, 1–16.

Karni, A., & Sagi, D. (1991). Where practice makes perfect. Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Science, USA, 88*, 4966–4970.

Kuefner, D., Jacques, C., Prieto, E. A., & Rossion, B. (2010). Electrophysiological correlates of the composite face illusion: Disentangling perceptual and decisional components of holistic face processing in the human brain. *Brain and Cognition, 74*, 225–238.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255–259.

Meinhardt, G., & Grabbe, Y. (2002). Attentional control in learning to discriminate bars and gratings. *Experimental Brain Research, 142*, 539–550.

Meinhardt-Injac, B., Persike, M., & Meinhardt, G. (2010). The time course of face matching by internal and external features: Effects of context and inversion. *Vision Research, 50*, 1598–1611.

Meinhardt-Injac, B., Persike, M., & Meinhardt, G. (2011). The time course of face matching for featural and relational image manipulations. *Acta Psychologica, 137*, 48–55.

Riesenhuber, M., & Wolff, B. S. (2009). Task effects, performance levels, features, configurations, and holistic face processing: A reply to Rossion. *Acta Psychologica, 132*, 286–292.

Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society of London B, 271*(Suppl.), S448–S450.

Richler, J. J., Tanaka, J. W., Brown, D. D., & Gauthier, I. (2008). Why does selective attention to parts fail in face processing? *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*, 1356–1368.

Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing happens at a glance. *Vision Research, 49*, 2856–2861.

Richler, J. J., Bukach, C. M., & Gauthier, I. (2009). Context influences holistic processing of nonface objects in the composite task. *Attention, Perception, & Psychophysics, 71*, 530–540.

Roelfsema, P. R., van Ooyen, A., & Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends in Cognitive Sciences, 14*, 64–71.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica, 128*, 274–289.

Rossion, B., & Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: Evidence from the face composite illusion. *Journal of Vision, 8*, 1–13.

Roufs, J. A. (1972). Dynamic properties of vision – I. Experimental relationships between flicker and flash threshold. *Vision Research, 12*, 261–278.

Roufs, J. A., & Blommaert, F. J. (1981). Temporal impulse and step responses of the human eye obtained psychophysically by means of a drift-correcting perturbation technique. *Vision Research, 21*, 1203–1221.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychological Science, 5*, 195–200.

Seitz, A. R., & Watanabe, T. (2003). Is subliminal learning really passive? *Nature, 422*, 36.

Seitz, A., & Watanabe, T. (2005). A unified model for perceptual learning. *Trends in Cognitive Science, 9*, 329–334.

Sekunova, A., & Barton, J. J. S. (2008). The effects of face inversion on the perception of long-range and local spatial relations in eye and mouth configuration. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 91129–91135.

Sergent, J. (1986). Microgenesis of face perception. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. M. Young (Eds.), *Aspects of face process information in recognition tasks*. Dordrecht: Martinus Nijhoff.

Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception & Psychophysics, 52*, 582–588.

Sinha, P., & Poggio, T. (1996). I think I know that face.... *Nature, 384*, 404.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400*, 869–873.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 45*, 34–79.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25*, 583–592.

Thomas, R. D. (2001). Perceptual interactions of facial dimensions in speeded classification and identification. *Perception & Psychophysics, 63*, 625–650.

Tovee, M. J., Rolls, E. T., Treves, A., & Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *Journal of Neurophysiology, 70*, 640–654.

Victor, J. D., & Shapley, R. M. (1979). Receptive field mechanisms of cat X and Y retinal ganglion cells. *The Journal of General Physiology, 74*, 275–298.

Williams, C. C., & Henderson, J. (2007). The face inversion effect is not a consequence of aberrant eye movements. *Memory & Cognition, 35*, 1977–1985.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*, 747–759.

Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron, 44*, 889–898.